DOCUMENT RESUME

ED 318 781 TM 014 885

AUTHOR Busch, John Christian
TITLE Application of a Mixed Consequential Ethical Model to a Problem Regarding Test Standards.
PUB DATE Apr 90
NOTE 22p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Viewpoints (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Criterion Referenced Tests; *Ethics; Licensing Examinations (Professions); *Models; Moral Values; Psychometrics; *Standards; *Teacher Certification; *Testing Problems; Value Judgment
IDENTIFIERS Curran (Charles); *Mixed Consequentialist Ethical Model; *Standard Setting

ABSTRACT

The work of the ethicist Charles Curran and the problem-solving strategy of the mixed consequentialist ethical model are applied to a traditional social science measurement problem--that of how to adjust a recommended standard in order to be fair to the test-taker and society. The focus is on criterion-referenced teacher certification tests. Because of test score fallibility, decisions made with those scores are subject to error, such as in the case of labeling a truly competent person as incompetent, or the reverse, labeling an incompetent person as competent. Psychometricians must recommend decisions regarding the standard; the ethical model of Charles Curran is adopted in this discussion. Curran specifically addresses the need for ethicists to speak to science in considering ethical issues, and he emphasizes the importance of an interchange between the technical and the theological/philosophical. A central and critical aspect of Curran's ethical model is the identification of the values of oneself or the institution. The consequences of adjusting the standard are to be examined and then compared with the values sought. Resolution then depends on the relative evaluation of the problems resulting from false positive and false negative errors. The advantages must outweigh the disadvantages. Values judged to be relevant to the problem in the case of teacher certification are listed, and a matrix summarizing consistency of consequences of actions with values is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED318781

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

JOHN CHRISTIAN BUSCH

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Application of a Mixed Consequential Ethical Model to a Problem Regarding Test Standards [1]

John Christian Busch
The University of North Carolina at Greensboro

Presented at the Annual Meeting of The American Educational Research Association, Boston MA, April 19, 1990

1. Supported in part by the Center for Educational Studies and Development, University of North Carolina at Greensboro

TM014985

BEST COPY AVAILABLE

I. Introduction

Context

Standardized testing is a pervasive experience in our culture; it would be nearly impossible today to find someone who has not taken at least one standardized test. In addition, the personal and collective effects of standardized testing are substantial. For example, testing plays a part in whether one is accepted for study at a university, whether one is accepted to a special educational program, whether one secures a job, whether one receives a promotion and whether one may practice a profession. The decision to use tests has significant human consequences. Curran (1985a, p. 55) discussing a different issue states that "...these decisions are more than merely technical decisions, they are truly human moral and Christian decisions. Christian ethics and the Church have something to say about those significant human issues". The same statement may be made with respect to educational testing.

For the past twenty years my scholarly interests have focused on applied measurement problems of a technical nature. My training and scholarship has been conducted within the traditions of the behavioral-social sciences. In the present study I attempted to utilize the methodology, questions and truth testing strategies of the ethicist to consider a traditional social science, measurement problem. Specifically, I have utilized the work of the ethicist, Charles Curran and the problem solving strategy of the mixed consequentialist ethical model to address a significant applied measurement problem.[1] The problem studied, deals with one aspect of standard setting for an educational test. Utilizing the methodology of ethics appears to be consistent with recent conceptualizations of test validity which recognize the need to assess values and to evaluate the consequences of actions based on testing (e.g., Messick, 1989). The current analysis is directly grounded in my on-going research on the technical aspects of standard-setting (e.g. Busch & Jaeger, in press, 1990; Busch & Jaeger, 1989).

Organization of the Paper

This paper is organized into four major parts. The second section provides background information for the reader on the substantive area of measurement and standard-setting and reviews some significant

theoretical writing on test validity which has important implications for this paper. The third section discusses the work of Charles Curran and identifies the components of a practical model for problem solving which is mixed consequential in nature. The fourth section utilizes this model to address a specific measurement issue.

Limitations and Assumptions

The measurement problem addressed in this paper focuses on a very limited and circumscribed technical issue. It should be stressed that the problem addressed in this paper is not how to obtain a recommended test standard, that issue is dealt with in the technical literature. Rather the issue is, once a recommended standard has been obtained, should I adjust it in order to be fair to the test-taker or to society? If so, how shall I adjust it? The problem is described in more detail below.

It is recognized that there are prior, broader and more significant questions that should also be asked. For example, should the test be given in the first place, and what are the ethical implications of differential test performance by various racial or gender groups? The narrower question addressed here has been selected because it is simple and will allow the investigator to evaluate the utility of the approach used by the ethicist to address questions that have been traditionally technical and scientific measurement issues. If this application to a circumscribed problem yields satisfactory results, then the methodology can be applied to more complex problems. This paper limits its focus to criterion referenced teacher certification tests.

The issue studied in the paper isn't limited to paper and pencil, standardized tests but potentially includes evaluations of many types, collected from a variety of sources. It should be acknowledged that the "standards" issue is not limited to standardized tests. We set standards of quality for many of our important activities e.g.. reviews of AERA paper proposals, grades for our students, for the books we read, for our family physicians etc. Finally, it should be stated that standardized tests themselves aren't good or bad in the abstract, however it is the use of a test for a particular purpose that may be morally good or bad.

II. Measurement Issues

A test standard or cutting score is a point on a score scale at or above which one passes a test. Legislation or policy statements frequently prescribe that groups of individuals must "pass" a particular certifying or licensing examination but do not operationally define exactly what "passing" means. Recommendations for a passing score are often obtained from panels of "expert" judges who are knowledgeable about the domain of knowledge or skills being measured by the test. There are a number of methodologies used to approach the standard-setting problem. The end result of the process is a specific score which defines the lowest acceptable passing score.

The problem addressed in his paper assumes that such a panel has made its recommendations and that the psychometrician has computed the panel's average recommended standard. However, because recommended test scores are fallible, i.e., they contain errors, one must decide whether to adjust that score. The possibilities for the psychometrician are to leave the recommended passing score where it is, or to raise or lower the recommended passing score. The problem addressed here is, can one utilize the methods of problem solving used by ethicists operating out of a mixed consequential model to solve this problem? Finally, what decision is made using that model?

Recent Analysis of Test Validity

Samuel Messick's chapter on test validity in the 3rd edition of Educational Measurement (1989) is a statement of the current state of knowledge of the psychometric community. His chapter is notable for its interpretations of validity issues which incorporate concepts which relate to ethical problem solving. Messick began discussing those issues as early as 1965 and has continued his intellectual commitment to the present. He (1975) identifies two basic measurement questions: The first asks whether a test is a good measure of the trait. This is a technical, scientific and a psychometric question. The second asks whether a test should be used for a particular purpose and is a question which asks about the ethics of a decision. The present study proposes to address a

measurement problem utilizing a particular ethical model to arrive at the decision.

The identification of value implications, and the need to evaluate the intended and unintended consequences of testing are incorporated in Messick's summary of the facets of test validity. He states that "The consequential basis of test interpretation is the appraisal of the value implications of the construct label, of the theory underlying test interpretation, and of the ideologies in which the theory is embedded... Finally, the consequential basis of test use is the appraisal of both potential and actual social consequences of the applied testing" (1989, p. 20).

The Role of Values in Test Validation

Messick (1989) recognizes that values must be addressed in all scientific endeavors including test validation efforts. At a fundamental level, values determine what is, and what is not measured. He states that "...the value implications of score interpretations should be explicitly addressed as part of the validation process itself" (1989, p. 59). They are important to validation because values influence (bias) interpretations made of the scores and of the construct. While bias can not be completely eliminated, science does require that biases be taken into account when interpretations are made. Howard (1985) states that, "the controversy is no longer about whether values influence scientific practice, but rather about how values are embedded in and shape scientific practice" (p. 255).

Messick (1980) gives several examples of the process of value identification in problems of test fairness. The identification of problems is not a panacea and he states that "...making values explicit does not determine choices among them, and at this point it appears difficult if not impossible..." (p. 1021) to balance the needs of a variety of audiences. We need methods to identify and analyze values associated with test validation. Messick (1989) states that "We need to explore some ways and means of uncovering tacit value premises and of coping with their consequences for test validation" (p. 62). He suggests a number of approaches including the comparative techniques and dialectic strategies of Kant, Hegel and Singer (1959). Churchman (1971) uses an approach which involves contrasting each construct theory with antithetical,

although plausible counterperspectives which can be used to interpret scores.

An illustration of the need for multiple perspectives in assessing values is illustrated by Messick's judgment (1982) that the National Academy of Sciences report on ability testing edited by Wigdor & Garner, 1982 was biased toward institutional uses of ability tests (for example, for selection purposes) as opposed to individual uses by the respondent (for example for diagnostic uses). The mixed consequantial method of analysis of ethical issues utilized later requires the explicit identification of values and the evaluation of the consequences of actions in terms of those values. In my judgment, there are only a limited number of analyses which undertake to identify values and to evaluate the consequences of actions associated with test use. In fact, Messick (1989) points out that the Standards for Educational and Psychological Testing (1985) stop short of addressing the values issue.

Intended and Unintended Consequences

Messick (1989) asserts that the social consequences of test use, both intended and unintended must be explicitly assessed as part of the validation process, and that it is not sufficient to provide factual information about a test. Rather, summaries should include "some exposition of the critical value contexts in which the facts are embedded and with a provisional accounting of the potential social consequences of alternative test uses" (p. 88). The generation of these consequences is not easily accomplished. Messick (1975) suggests utilizing both the tester and the respondent's points of view in order to generate the antagonistic proposals discussed by Churchman (1968). Messick argues that multiple value perspectives should be used "to assure that a broad range of potential consequences of the testing are addressed in the validation process" (1989, p. 87). The issue of consequences is accounted for in the mixed consequential ethical model. In the model, the ethicist must deduce the likely consequences of various courses of actions and use that as a basis for decision making.

III. The Ethical Problem Solving Model

Why Charles Curran?

The work of Curran was selected for the following reasons. He specifically addresses the need for ethicists to speak to science in considering ethical issues and he stresses the importance of an interchange between the technical and the theological/philosophical. "Often Churchmen can be accused of "do goodism" because they betray an ignorance of the technical and factual data involved. Expertise is absolutely essential before such a judgment can be made. Obviously such knowledge calls for an extensive dialogue within the Church with those who have the technical and factual data" (1974, p. 155). He indicates that "Moral theology necessarily is in dialogue with the human and social sciences such as anthropology, economics, psychology and sociology. A very significant methodological question concerns the way in which moral theology integrates and uses the empirical sciences" (1982, p. 64).

He also states that theology has no special "insight into discerning the social problems facing the world and the church today". Theology "must learn from all possible sources what are the primary issues...must call for the cooperation of all in discerning these problems-especially minorities and those who are oppressed and suffering" (1985a, p. 106)

He recognizes that some of our ethical conclusions may change over time as we learn about various phenomena and he recognizes the decision making process as complex and on-going. He emphasizes the primacy of reason and responsibility. Finally his relationality-responsibility model is appealing; it is described below.

Where does Curran fit?

Gula (1982) and Curran (1988) contrast two world views within moral theology: the classical which is based on concepts of static, unchanging essences versus a more modern, historically conscious and dynamic perspective. The more modern world view is more responsive to our evolving understanding of human relationships. It is more inductive, and it emphasizes the need for change. It acknowledges the existence of errors of judgment and incorporates mechanisms for self correction. Such

a historically conscious method does not pretend to have absolute certainty (Curran,1985b).

Curran (1988) identifies three fundamental general ethical models: the deontological, the teleological and a relationality-responsibility model. In the deontological model, morality is considered in terms of law and its obligations (duty). Certain acts are intrinsically and always good or evil by their nature. As a result, that stance suggests the presence of certain absolute moral norms and it is associated with a classical world-view. In the teleological approach, the ethical life is viewed in terms of the goal to be achieved by actions. An action is judged good if it is conducive to achieving a desired end. Curran describes the third model, which he adopts as his own, as a relationality-responsibility model This model views the moral life in terms of a person's multiple relationships with God, neighbor, world, and self. Curran (1988) believes that one model will be primary but not exclusive. That is, a relationality-responsibility model could incorporate law within itself but law would not dominate

All consequential theories are teleological in nature. Consequential theories focus on the results of acts. Curran (1979) summarized contemporary objections to strict consequentialism to include the following: "(1) aspects other than consequences must be considered, (2) The good can't be determined independent of the morally right, (3) Not only the consequences of the action but also the way in which the action brings about the consequences has moral significance" (p. 348).

His relationality-responsibility model is not strictly consequential but considers a variety of factors necessary to come to a rational decision. The factors include attitudes, values, dispositions, intentions etc. "Questions of attitudes, dispositions, ideals, values, goals, perspectives and intentionalities have not received the attention they deserve in a complete consideration of moral theology" (1979, p. 355). He argues that "the relationality approach is not merely a middle approach between the other two but in a sense also opts for a somewhat different understanding of the moral decision making process" (1979, p. 357).

Curran (1979) describes himself as a mixed consequentialist. He draws a distinction between two systems of related ideas. The first system is the general model for the moral life, while the second applies to more practical methods for decision-making in concrete cases.

Regarding the first issue, he states "I prefer to accept an ethical model of relationality and responsibility as a third model distinguished from both teleology and deontology" (1979, p. 356). However, "On the level of the formulation of moral norms and the criteria for concrete decision making I would opt for the second position described above as mixed consequentialism" (1979, p. 356). A mixed consequential approach to decision making will be used in this paper.

Curran addresses the problem of weighing good and evil outcomes. The principle of proportionality is used to decide on a morally justified action when actions have mixed consequences. He states "The principle of proportionality maintains that the advantages that will be achieved by using these means must not be disproportionate to the harms reasonably expected to follow from use" (1988, p. 140).

One of the differences between Curran's position and a strictly consequential one is the acknowledgment of norms and the use of communal wisdom in arriving at an ethical decision. However, regarding the use of norms, Curran (1988) indicates that "moral norms in my judgment are not the primary, or the only, or the most concern of moral teaching and of moral theology". Rather, he indicates that "Values, attitudes, and dispositions are much more important and far reaching for the moral life than are specific norms" (p. 21).

A mixed consequential analysis (Gula, 1982) will be used as a methodology to address the measurement problem and will include the following components:

1. Identification of the problem.
2. Analysis of the problem.
    a. Who and what are involved in the problem.
    b. The context (when, where, why and how) of the problem.
3. Analysis of the relevant values.
    a. Perceptions of the values as influenced by one's beliefs and convictions.
    b. Identification of the values at stake in the situation.
4. Appeal to norms to guide actions which protect values.
    a. Identification and analysis of norms to identify possible actions.
    b. Exploration of the consequences of actions.

5. Compare consequences with values sought
    a. If consequences are inconsistent with values, then explore other alternatives and test alternatives by appeal to norms.
    b. If consequences are consistent with values, determine if the action is proportionate (the means used to to realize the value contributes to more good than harm), then conclude that it is a viable moral option
6. Apply test of generalizability.
    a. Ask what if everyone in a similar situation were to act this way?
7. Measure against the identity, integrity, and capacity of agent.

IV. Analysis of the Problem Using a Mixed Consequential Strategy

1. Identification of the Problem:

All psychological and educational measurement is fallible, there are errors of measurement. Because of this error it is possible that someone whose "true" test score was exactly equal to the standard (passing score) would obtain a test score which is greater or less than her true score. For example a person with true score actually equal to 30 might guess correctly on one item and have an obtained test score equal to 31. Or equally possible, the same individual might miss-mark a single item which s(he) knew and score 29. Therefore, test scores are approximations of an individuals true knowledge

A test standard is a passing score or cut score. Legislation or policy statements frequently prescribe that groups of individuals must pass a particular certifying or licensing examination but do not operationally define exactly what "passing" means. Recommendations for a passing score are often obtained from panels of "expert" judges who are knowledgeable about the domain being measured by the test. For example panels of public school history teachers, historians, teacher educators in history programs, curriculum experts in social studies, etc., might be involved in the judgment process for an achievement test in history. Panels would be representative of the various constituencies with an interest in the testing program. There are a number of methodologies used to obtain the recommended standards however no one method is

unanimously preferred. The end result of the process is a specific score which defines the lowest acceptable passing score. It is a fallible index i.e., it contains error of at least two types: measurement and sampling error. The issue addressed in this paper revolves around choices in dealing with that error.

2. Analysis of the Problem

a. Who and What are Involved in the Problem:

Because of test score fallibility, decisions made with those scores are subject to two types of errors of misclassification. The first is to call someone who is truly competent, incompetent (a "false negative" error). The second is to incorrectly call a truly incompetent person, competent (a "false positive" error). The test user, probably with the recommendation of a psychometric consultant must choose between three alternatives: keep the passing score as is, or to adjust it for the presence of the errors noted above. Those adjustments are either to set the passing score higher (by so many standard deviation units) or to lower it (by so many standard deviation units). Each of the latter two alternatives has distinct and competing consequences. If one sets the passing score above the minimum recommended level, the number of false positives is reduced, at the same time such an adjustment would result in a greater number of false negatives. On the other hand if the standard was adjusted downward, there would be fewer false negatives but more false positives. It seems to me that the higher standard protects society's interests and potentially sacrifices those interests of the qualified individual applicants while the lower standard protects the qualified individual applicant's while at the same time potentially sacrifices some of society's interest. The psychometrician must recommend a decision regarding the standard; she adopts the ethical model of Charles Curran to evaluate the rightness of the decision.

b. The Context (When, Where, Why and How) of the Problem:

This situation is one in which a criterion-referenced test is being used to assess the level of knowledge of a professional, for example a teacher who is applying for initial licensing or certification. All individuals seeking certification are required to pass this examination.

The test may be retaken an unlimited number of times. The professional must pass the test a single time in order to receive permanent certification. The test user is an agent of the state which has legal responsibility for administering educational policy. The test has not been used before, or if it has, the standard is being re-set. The issue which is being focused on is but one of a number which would be considered before making a decision to implement a testing program. That is issues of reliability, bias, test development etc would be evaluated individually. Some collective decision would have to be made about the overall use of the test based on the collection of individual decisions. It it also assumed that in-class teaching skills, in addition to knowledge would be important and that other ways of assessment would be used in the assessment process e.g., observation, case study, portfolio etc. It is assumed that subject matter knowledge is a necessary but not sufficient condition for competent teaching.

3. Analysis of the Relevant Values

a. Identification of the Values at Stake Based on One's Perceptions and Beliefs

A central and critical aspect of the ethical evaluation of the particular use of a test is the identification and specification of one's, or of the institution's values. While psychometricians have of late acknowledged that the uses of test scores are related to one's value system, I am unaware of no more than a few situations in which a clear statement of values has been obtained as a part of the preparation for test use. Exhibit 1 contains a list of my values which pertain to the particular problem addressed. The list is far from exhaustive. My future research

Insert Exhibit 1 about here

agenda will include study of methods for identifying and clarifying the values of test users, policy makers, test developers etc.

4. Appeal to Norms to Guide Actions Which Protects Values (explore consequences of action)

There are three actions that might be taken:
action 1: do not adjust the standard;
action 2: adjust the standard by lowering it;
action 3: adjust the standard by raising it.

a. Norms of behavior:

There have been very few cases of an unadjusted-standard being used (action 1). In most cases adjustments have lowered the standard (action 2). I know of no instances in which the standard has been raised (action 3). These are norms from the professional community.

b. Consequences of Actions:

The consequences of action 1 (using an unadjusted standard) are that 50% of those <u>just</u> at the minimal level of competence would fail the test, solely by chance (false negatives); and 50% would pass.

The consequence of lowering the standard (action 2) is to decrease the proportion of truly qualified individuals who would have failed the test due to measurement error from 50% of those just at the standard to a much smaller percent (false negatives). At the same time the proportion of unqualified individuals passing (false positives) will increase.

The consequence of raising the standard (action 3) would be to reduce the proportion of truly unqualified applicants who would have passed the test solely by chance (false positives). At the same time the percentage of truly qualified individuals failing the test will decrease.

5. Compare the Consequences of the Actions with the Values Sought

I have created a matrix which summarizes the results of that analysis. The consequences of the three actions is evaluated with respect to each of the values and judged to be consistent or inconsistent. Several of the values are basically neutral with respect to the consequences; in such cases I have categorized them as consistent. The matrix is contained in Exhibit 2.

Insert Exhibit 2 about here

The question seems to turn on the values best served by each of the choices. The values which discriminate the choices are as follows. If one values the reduction of decision errors of either kind (see value 7, in **bold**) then one would more likely choose actions #2 or #3. Second, values #9,l,10, 13,14, and 16 discriminate between actions 2 and 3. That is if one values the creation of a larger and more diverse pool of applicants, if one defers more to the rights of the individual teacher and one prefers to err in accepting rather than rejecting, then the strategy of adjusting the standard downward is most consistent with those values.

There are a mixture of good and bad outcomes depending on the choice of actions. Resolution of the problem depends upon one's relative evaluations of the harm which follows from making false positive and false negative errors. In whatever choice is made, there would have to be a greater proportion of good to evil.

My choice would be for action 2. My reasoning is as follows. There is first the matter of collective wisdom and norms which are not decisive but which would be significant elements in the solution. The collective wisdom of others who have studied this as a <u>technical</u> issue has almost always favored a downward adjustment of the standard. From a western cultural point of view, I believe that we have traditionally attended to the rights of the individual when in potential conflict with the collective good.

Which is more serious with respect to teachers seeking certification, a false positive or a false negative error? I would judge that a false negative decision is the more serious error and should be reduced even at the cost of increased false positive error. In short, my valuation of the individual (the teacher), and of diversity and pluralism leads me to that choice. I prefer to make an error of acceptance rather than to err in rejecting. I believe that the harm done in failing an individual teacher is certain and immediate, while the possible harm done in passing an incompetent teacher is more distant and contingent upon other factors. For example, it is possible that a truly incompetent teacher who passed the test would study, or otherwise augment her knowledge

through course work etc., thus lessening the potential harm to society (individual students).

The proportionality theory states that the good achieved must outweigh the negative outcomes. One must weigh the good of decreasing the number of misclassifications of qualified teachers against tolerating the evil of admitting some individuals to the teaching profession who may not be qualified. Gula suggests the following tests: It is assumed that the value (#10, 13, 14) supported is at least equal to that which is sacrificed (#1, 9, 16); that there is no less harmful way to protect the value here and now; the manner of protection here and now will not undermine it in the long-run. I believe that the proportionality criterion is satisfied in this instance.

6. Apply Test of Generalizability

If everyone were to adopt this decision, I believe the result might be satisfactory in similar testing situations. I would stop short of proscribing the decision as a general solution, that is that it is truly generalizable. It must be noted that a number of parameters would influence such a decision. For example, would the decision be the same if the test were used to screen the competence of physicians? Would the thoroughness of the standard-setting method, as practiced in a particular situation, influence the decision? What would the effect be if a construct less well understood were being tested, such as those aspects of teaching effectiveness as empathy, ability to communicate abstract ideas, ability to promote critical thinking skills. Certainly, as our knowledge of the teaching process evolves the sorts of decisions we must make will change and will be dependent upon our knowledge base in the social sciences.

7. Measure Against the Identity, Integrity and Capacity of the Agent

The decision which is made here is subject to a great deal of uncertainty. The decision maker can not identify all relevant values, nor can she comfortably deduce all possible consequences. Therefore, such conclusions must be regarded as the best decision for the present. There must be an openness to change and the evolution of different perspectives over time. The decision maker must actively seek the participation of different groups in playing out those alternative perspectives.

This analysis has been conducted assuming that the agent is a psychometric expert who is a technically competent and experienced professional. While such a person may be presumed to be of good will, we must always recognize that the scholarly socialization processes produce individuals with relatively homogeneous world views and perhaps values. Additionally, it should be recognized that the decision maker in situations such as this is often not the psychometrician but a policy maker such as a member of a State Board of Education, etc. Nevertheless, appropriate recommendations by the technical expert can influence the policy maker to some extent.

References

Busch, J. C. & Jaeger, R. M. (In press, Summer, 1990). Influence of type of judge, normative information and discussion on standards recommended for the National Teacher Examinations. Journal of Educational Measurement.

Busch, J. C. & Jaeger, R. M. (1989). Policy and Educational Standards. Administrator's Notebook.

Churchmann, C. W. (1968). The systems approach. New York: Delacorte Press.

Churchmann, C.W. (1971). The design of inquiring systems: Basic concepts of systems and organization. New York, NY: Basic Books

Curran, C. (1974). New perspectives in moral theology. Notre Dame University, IN: Fides Press.

Curran, C. (1979). Utilitarianism and contemporary moral theology: Situating the debates, In C. Curran & R. McCormick (Eds.), Readings in moral theology: No. 1, New York, NY: Paulist Press.

Curran, C. (1982). Moral theology: A continuing journey. Notre Dame University, IN: University of Notre Dame Press.

Curran, C. (1985a). Directions in Catholic social ethics. Notre Dame University, IN: University of Notre Dame Press.

Curran, C. (1985b). Directions in fundamental moral theology. Notre Dame, IN.:University of Notre Dame.

Curran, C. (1987). Toward an american catholic moral theology. Notre Dame University, IN: University of Notre Dame Press.

Curran, C. (1988). Tensions in moral theology. University of Notre Dame, IN: University of Notre Dame Press.

Gula, R. (1982). What are they saying about moral norms? New York: Paulist Press.

Howard, G.S. (1985). The role of values in the science of psychology. American Psychologist, 40(3), 255-265.

Messick, S. A. (1965). Personality measurement and the ethics of assessment. American Psychologist, 20, 136-142.

Messick, S. A. (1975). The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 30, 955-966.

Messick, S. A. (1980). Test validity and the ethics of assessment. American Psychologist, 35(11), 1012-1027.

Messick, S.A. (1982). The values of ability testing: Implications of multiple perspectives about criteria and standards. Educational Measurement: Issues and Practice, 1(3), 9-12, 20, 26.

Messick, S. (1989). Validity. In Linn, R. (editor), Educational Measurement, (3rd edition), Washington, DC: American Council on Education/American Educational Research Association.

Shapere, D. (1977). Scientific theories and their domain. In F. Suppe (Ed.), The structure of scientific theories (2nd edition, p. 518-589) Urbana, Ill, University of Illinois Press.

Singer, E.A. (1959). Experience and reflection. (C.W. Churchman, Ed.), Philadelphia: University of Pennsylvania Press.

Wigdor, A. K. & Garner, W.R. (Eds.).(1982). Ability testing: Uses, consequences, and controversies. Washington, DC: National Academy Press.

Author Notes

1. Work on this paper was begun when the author was a fellow at the Coolidge Research Colloquium, Cambridge, MA, June 1989.

2. I wish to thank the following colleagues for their thoughtful comments and suggestions: various 1989 Coolidge fellows especially Professors Carol Ann Wildt and Ann Stankewitz and my colleagues at UNCG, Patrick Hensy, David Pratto and David Purpel.

Exhibit 1

List of Values Judged to be Relevant to the Problem Situation

Value 1: Incompetent teachers should not be allowed to teach our children. (quality education is valued)

Value 2: Test scores are fallible; therefore they should be treated as approximations of a person's "true" level of knowledge.

Value 3: Although fallible, scores obtained from tests may be acceptably valid and reliable indicators of a student's status with respect to a particular construct. (test scores are useful)

Value 4: As human creatures created by God, all individuals are entitled to and should be treated with dignity. (human dignity is valued)

Value 5: Individuals should be treated justly. (justice to individuals is valued)

Value 6: Distributive justice; All individuals have a right to a fair share of the goods of this world (Certification involves the right to earn a living).

Value 7: Decision errors from a fallible test are unavoidable and should be minimized. (Statistical precision is valued)

Value 8: Standards should be estimated by panels of informed representative judges. (Expertise by judges is valued)

Value 9: Knowledgeable teachers are valued more than less knowledgeable teachers. (Teacher competence is valued)

Value 10: The rights of the individual teacher must be given respect. (Individual teachers' rights are valued)

Value 11: Objective measurement is preferred to subjective measurement. (Objectivity valued)

Value 12 Individuals should treat each other with respect and love. (Love of Neighbor valued)

Value 13: I would rather make a false positive error than a false negative; I would rather make the mistake of calling someone competent who is not. (Greater tolerance for the first error)

Value 14: Preference should be given to procedures which produce a more heterogeneous pool of applicants. (Heterogeneity valued)

Value 15: There is no certitude in our decision making; a certain amount of in-exactness is tolerable. (Tolerance in the face of a fallible system is valued)

Value 16: The rights of society (students in an educational system, taken collectively and their parents) to a satisfactory education should be supported. (Quality education is valued

Exhibit 2

Matrix Summarizing Consistency of Consequences of Actions with Values

ACTION	CONSEQUENCE	CONSISTENT WITH VALUES:	INCONSISTENT WITH VALUES:
1 (no adjust)	no change in errors	2,3,4,5,6,8,9,11,12	1,7,10,13 14,15,16
2 (lower)	false neg decrease false pos increase	2,3,4,5,6,7,8, 10,11,12,13,14,15	1, 9,16
3 (increase)	false pos decrease false neg increase	1,2,3,4,5,6,7,8,9, 11,12,15,16	10,13,14